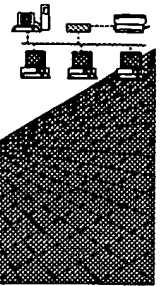Roner

RAW SEQUENCE LISTING
ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following CRF diskette:

Application Serial Number: 09/051,670

Art Unit / Team No. : 1646

Date Processed by STIC: 12/21/98

THE ATTACHED PRINTOUT EXPLAINS THE **ERRORS DETECTED.**

PLEASE BE SURE TO FORWARD THIS INFORMATION TO THE APPLICANTS BY EITHER:

1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANTS ALONG WITH A NOTICE TO COMPLY or,

2) CALLING APPLICANTS AND FAXING THEM A COPY OF THE PRINTOUT WITH A NOTICE TO COMPLY

THIS WILL INSURE THAT THE NEXT SUBMISSION RECEIVED FROM THEM WILL BE ERROR FREE.

IF YOU HAVE ANY FURTHER QUESTIONS, PLEASE CALL:

ARTI SHAH 703-308-4212

Raw Sequence Listing Error Summary

ERROR DETECTED SUGGESTED CORRECTION   SERIAL NUMBER: _09/051,670_

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

1 _____ Wrapped Nucleics   The number/text at the end of each line "wrapped" down to the next line.
                               This may occur if your file was retrieved in a word processor after creating it.
                               Please adjust your right margin to .3, as this will prevent "wrapping".

2 _____ Wrapped Aminos   The amino acid number/text at the end of each line "wrapped " down to the next line.
                               This may occur if your file was retrieved in a word processor after creating it.
                               Please adjust your right margin to .3, as this will prevent "wrapping".

3 __√__ Incorrect Line Length The rules require that a line not exceed 72 characters in length. This includes spaces.
                               All text must be visible on page.

4 _____ Misaligned Amino Acid . The numbering under each 5th amino acid is misaligned. This may be caused by the use of tabs
        Numbering   between the numbering. It is recommended to delete any tabs and uses spacing between the numbers.

5 _____ Non-ASCII   This file was not saved in ASCII (DOS) text, as required by the Sequence Rules.
                               Please ensure your subsequent submission is saved in ASCII text so that it can be processed.

6 _____ Variable Length   Sequence(s) _____ contain n's or Xaa's which represented more than one residue.
                               As per the rules, each n or Xaa can only represent a single residue.
                               Please present the maximum number of each residue having variable length and
                               indicate in the (ix) features section that some may be missing.

7 _____ Wrong Designation   Sequence(s) _____ contain amino acid or nucleic acid designators which are not standard
                               representations as per the Sequence Rules (Please refer to paragraph 1.822)

8 _____ Skipped Sequences   Sequence(s) _____ missing. If intentional, please use the following format for each skipped sequence:
        (OLD RULES)   (2) INFORMATION FOR SEQ ID NO:X:
                               (I) SEQUENCE CHARACTERISTICS:(Do not insert any headings under "SEQUENCE CHARACTERISTICS")
                               (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X:
                               This sequence is intentionally skipped

                               Please also adjust the "(iii) NUMBER OF SEQUENCES:" response to include the skipped sequence(s).

9 _____ Skipped Sequences   Sequence(s) _____ missing. If intentional, please use the following format for each skipped sequence.
        (NEW RULES)   <210> sequence id number
                               <400> sequence id number
                               000

10 _____ Use of N's or Xaa's   Use of N's and/or Xaa's have been detected in the Sequence Listing.
        (NEW RULES)   Use of <220> to <223> is MANDATORY if n's or Xaa's are present.

11 _____ Use of <213>Organism Sequence(s) _____ are missing this mandatory field or its response.
        (NEW RULES)

12 _____ Use of <220>Feature   Sequence(s) _____ are missing the <220>Feature and associated headings.
        (NEW RULES)   Use of <220> to <223> is MANDATORY if <213>ORGANISM is "Artificial" or "Unknown"
                               (See "Federal Register," 6/01/98, Vol. 63, No. 104, pp. 29631-32)
                               (Sec. 1.823 of new Sequence Rules)

13 _____ Wrong Format   File submitted was in the alphabetical heading format of the Old Sequence Rules.This is invalid since the
                               "Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Disclosures"
                               Federal Register Notice, Vol. 63, No. 104, June 1, 1998, p. 29620
                               applies to applications filed on or after July 1, 1998.

SEQUENCE LISTING 09/051,610

(1) GENERAL INFORMATION:
Sequence list
(i) Applicant: SNOW BRAND MILK PRODUCTS CO.,
(ii) Title of the Invention: Novel DNA and process for preparing protein using the DN
(iii) Number of Sequences: 4

(iv) CORRESPONDENCE ADDRESS:
 (A) ADDRESSEE:
 (B) STREET:
 (C) CITY:
 (D) STATE:
 (E) COUNTRY:
 (F) ZIP:
(v) COMPUTER READABLE FORM:
 (A) MEDIUM TYPE:
 (B) COMPUTER:
 (C) OPERATING SYSTEM:
 (D) SOFTWARE:
(vi) CURRENT APPLICATION DATA:
 (A) APPLICATION NUMBER:
 (B) FILING DATE:
 (C) CLASSIFICATION:

(2) INFORMATION FOR SEQ ID NO: 1:
Sequence number: 1
(A) Length of sequence: 1316
(B) Sequence type: nucleic acid
(C) Strandedness: double
(D) Topology: linear
(ii) Molecule type: genomic DNA (......ic DNA-1)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
Sequence:

(i) SEQUENCE CHARACTERISTICS:

```
CTGGAGACAT ATAACTTGAA CACTTGGCCC TGATGGGGAA GCAGCTCTGC AGGGACTTTT    60
TCAGCCATCT GTAAACAATT TCAGTGGCAA CCCGCGAACT GTAATCCATG AATGGGACCA   120
CACTTTACAA GTCATCAAGT CTAACTTCTA GACCAGGGAA TTAATGGGGG AGACAGCGAA   180
CCCTAGAGCA AAGTGCCAAA CTTCTGTCGA TAGCTTGAGG CTAGTGGAAA GACCTCGAGG   240
AGGCTACTCC AGAAGTTCAG CGCGTAGGAA GCTCCGATAC CAATAGCCCT TTGATGATGG   300
TGGGGTTGGT GAAGGGAACA GTGCTCCGCA AGGTTATCCC TGCCCCAGGC AGTCCAATTT   360
TCACTCTGCA GATTCTCTCT GGCTCTAACT ACCCCAGATA ACAAGGAGTG AATGCAGAAT   420
AGCACGGGCT TTAGGGCCAA TCAGACATTA GTTAGAAAAA TTCCTACTAC ATGGTTTATG   480
TAAACTTGAA GATGAATGAT TGCGAACTCC CCGAAAAGGG CTCAGACAAT GCCATGCATA   540
AAGAGGGGCC CTGTAATTTG AGGTTTCAGA ACCCGAAGTG AAGGGGTCAG GCAGCCGGGT   600
ACGGCGGAAA CTCACAGCTT TCGCCCAGCG AGAGGACAAA GGTCTGGGAC ACACTCCAAC   660
TGCGTCCGGA TCTTGGCTGG ATCGGACTCT CAGGGTGGAG GAGACACAAG CACAGCAGCT   720
GCCCAGCGTG TGCCCAGCCC TCCCACCGCT GGTCCCGGCT GCCAGGAGGC TGGCCGCTGG   780
CGGGAAGGGG CCGGGAAACC TCAGAGCCCC GCGGAGACAG CAGCCGCCTT GTTCCTCAGC   840
CCGGTGGCTT TTTTTTCCCC TGCTCTCCCA GGGGACAGAC ACCACCGCCC CACCCCTCAC   900
GCCCCACCTC CCTGGGGGAT CCTTTCCGCC CCAGCCCTGA AAGCGTTAAT CCTGGAGCTT   960
TCTGCACACC CCCCGACCGC TCCCGCCCAA GCTTCCTAAA AAAGAAAGGT GCAAAGTTTG  1020
GTCCAGGATA GAAAAATGAC TGATCAAAGG CAGGCGATAC TTCCTGTTGC CGGGACGCTA  1080
TATATAACGT GATGAGCGCA CGGGCTGCGG AGACGCACCG GAGCGCTCGC CCAGCCGCCG  1140
CCTCCAAGCC CCTGAGGTTT CGGGGGACCA CA ATG AAC AAG TTG CTG TGC TGC     1193
                                       Met Asn Lys Leu Leu Cys Cys
                                        -20                   -15
```

```
GCG CTC GTG GTAAGTCCCT GGGCCAGCCG ACGGGTGCCC GGCGCCTGGG           1242
Ala Leu Val
```

```
GAGGCTGCTG CCACCTGGTC TCCCAACCTC CCAGCGGACC GGCGGGGAGA AGGCTCCACT  1302
CGCTCCCTCC CAGG                                                   1316
```

(2) INFORMATION FOR SEQ ID NO: 2:
Sequence number: 2
(A) Length of sequence: 9898
(B) Sequence type: nucleic acid
(C) Strandedness: double
(D) Topology: linear
(ii) Molecule type: genomic DNA (human OCIF genomic DNA-2)
Sequence:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
GCTTACTTTG TGCCAAATCT CATTAGGCTT AAGGTAATAC AGGACTTTGA GTCAAATGAT    60
ACTGTTGCAC ATAAGAACAA ACCTATTTTC ATGCTAAGAT GATGCCACTG TGTTCCTTTC   120
TCCTTCTAG TTT CTG GAC ATC TCC ATT AAG TGG ACC ACC CAG GAA ACG TTT  171
            Phe Leu Asp Ile Ser Ile Lys Trp Thr Thr Gln Glu Thr Phe
                 -10                  -5                   1
```

```
CCT CCA AAG TAC CTT CAT TAT GAC GAA GAA ACC TCT CAT CAG CTG TTG    219
Pro Pro Lys Tyr Leu His Tyr Asp Glu Glu Thr Ser His Gln Leu Leu
```

```
                                  10                        15
TGT GAC AAA TGT CCT CCT GGT ACC TAC CTA AAA CAA CAC TGT ACA GCA     267
Cys Asp Lys Cys Pro Pro Gly Thr Tyr Leu Lys Gln His Cys Thr Ala
20            25                  30                  35

AAG TGG AAG ACC GTG TGC GCC CCT TGC CCT GAC CAC TAC TAC ACA GAC     315
Lys Trp Lys Thr Val Cys Ala Pro Cys Pro Asp His Tyr Tyr Thr Asp
                40                  45                  50

AGC TGG CAC ACC AGT GAC GAG TGT CTA TAC TGC AGC CCC GTG TGC AAG     363
Ser Trp His Thr Ser Asp Glu Cys Leu Tyr Cys Ser Pro Val Cys Lys
              55                60                  65

GAG CTG CAG TAC GTC AAG CAG GAG TGC AAT CGC ACC CAC AAC CGC GTG     411
Glu Leu Gln Tyr Val Lys Gln Glu Cys Asn Arg Thr His Asn Arg Val
          70                  75                  80

TGC GAA TGC AAG GAA GGG CGC TAC CTT GAG ATA GAG TTC TGC TTG AAA     459
Cys Glu Cys Lys Glu Gly Arg Tyr Leu Glu Ile Glu Phe Cys Leu Lys
      85                  90                  95

CAT AGG AGC TGC CCT CCT GGA TTT GGA GTG GTG CAA GCT G GTACGTGTCA    509
His Arg Ser Cys Pro Pro Gly Phe Gly Val Val Gln Ala
100             105             110

ATGTGCAGCA AAATTAATTA GGATCATGCA AAGTCAGATA GTTGTGACAG TTTAGGAGAA    569
CACTTTTGTT CTGATGACAT TATAGGATAG CAAATTGCAA AGGTAATGAA ACCTGCCAGG    629
TAGGTACTAT GTGTCTGGAG TGCTTCCAAA GGACCATTGC TCAGAGGAAT ACTTTGCCAC    689
TACAGGGCAA TTTAATGACA AATCTCAAAT GCAGCAAATT ATTCTCTCAT GAGATGCATG    749
ATGGTTTTTT TTTTTTTTTT TAAAGAAACA AACTCAAGTT GCACTATTGA TAGTTGATCT    809
ATACCTCTAT ATTTCACTTC AGCATGGACA CCTTCAAACT GCAGCACTTT TTGACAAACA    869
TCAGAAATGT TAATTTATAC CAAGAGAGTA ATTATGCTCA TATTAATGAG ACTCTGGAGT    929
GCTAACAATA AGCAGTTATA ATTAATTATG TAAAAAATGA GAATGGTGAG GGGAATTGCA    989
TTTCATTATT AAAAACAAGG CTAGTTCTTC CTTTAGCATG GGAGCTGAGT GTTTGGGAGG   1049
GTAAGGACTA TAGCAGAATC TCTTCAATGA GCTTATTCTT TATCTTAGAC AAAACAGATT   1109
GTCAAGCCAA GAGCAAGCAC TTGCCTATAA ACCAAGTGCT TTCTCTTTTG CATTTTGAAC   1169
AGCATTGGTC AGGGCTCATG TGTATTGAAT CTTTTAAACC AGTAACCCAC GTTTTTTTTC   1229
TGCCACATTT GCGAAGCTTC AGTGCAGCCT ATAACTTTTC ATAGCTTGAG AAAATTAAGA   1289
GTATCCACTT ACTTAGATGG AAGAAGTAAT CAGTATAGAT TCTGATGACT CAGTTTGAAG   1349
CAGTGTTTCT CAACTGAAGC CCTGCTGATA TTTTAAGAAA TATCTGGATT CCTAGGCTGG   1409
ACTCCTTTTT GTGGGCAGCT GTCCTGCGCA TTGTAGAATT TTGGCAGCAC CCCTGGACTC   1469
TAGCCACTAG ATACCAATAG CAGTCCTTCC CCCATGTGAC AGCCAAAAAT GTCTTCAGAC   1529
ACTGTCAAAT GTCGCCAGGT GGCAAAATCA CTCCTGGTTG AGAACAGGGT CATCAATGCT   1589
AAGTATCTGT AACTATTTTA ACTCTCAAAA CTTGTGATAT ACAAAGTCTA AATTATTAGA   1649
CGACCAATAC TTTAGGTTTA AAGGCATACA AATGAAACAT TCAAAAATCA AATCTATTC    1709
TGTTTCTCAA ATAGTGAATC TTATAAAATT AATCACAGAA GATGCAAATT GCATCAGAGT   1769
CCCTTAAAAT TCCTCTTCGT ATGAGTATTT GAGGGAGGAA TTGGTGATAG TTCCTACTTT   1829
CTATTGGATG GTACTTTGAG ACTCAAAAGC TAAGCTAAGT TGTGTGTGTG TCAGGGTGCG   1889
GGGTGTGGAA TCCCATCAGA TAAAAGCAAA TCCATGTAAT TCATTCAGTA AGTTGTATAT   1949
GTAGAAAAAT GAAAAGTGGG CTATGCAGCT TGGAAACTAG AGAATTTTGA AAAATAATGG   2009
AAATCACAAG GATCTTTCTT AAATAAGTAA GAAAATCTGT TTGTAGAATG AAGCAAGCAG   2069
GCAGCCAGAA GACTCAGAAC AAAAGTACAC ATTTTACTCT GTGTACACTG GCAGCACAGT   2129
GGGATTTATT TACCTCTCCC TCCCTAAAAA CCCACACAGC GGTTCCTCTT GGGAAATAAG   2189
AGGTTTCCAG CCCAAAGAGA AGGAAAGACT ATGTGGTGTT ACTCTAAAAA GTATTTAATA   2249
ACCGTTTTGT TGTTGCTGTT GCTGTTTTGA AATCAGATTG TCTCCTCTCC ATATTTTATT   2309
TACTTCATTC TGTTAATTCC TGTGGAATTA CTTAGAGCAA GCATGGTGAA TTCTCAACTG   2369
```

```
TAAAGCCAAA TTTCTCCATC ATTATAATTT CACATTTTGC CTGGCAGGTT ATAATTTTTA 2429
TATTTCCACT GATAGTAATA AGGTAAAATC ATTACTTAGA TGGATAGATC TTTTTCATAA 2489
AAAGTACCAT CAGTTATAGA GGGAAGTCAT GTTCATGTTC AGGAAGGTCA TTAGATAAAG 2549
CTTCTGAATA TATTATGAAA CATTAGTTCT GTCATTCTTA GATTCTTTTT GTTAAATAAC 2609
TTTAAAAGCT AACTTACCTA AAAGAAATAT CTGACACATA TGAACTTCTC ATTAGGATGC 2669
AGGAGAAGAC CCAAGCCACA GATATGTATC TGAAGAATGA ACAAGATTCT TAGGCCCGGC 2729
ACGGTGGCTC ACATCTGTAA TCTCAAGAGT TTGAGAGGTC AAGGCGGGCA GATCACCTGA 2789
GGTCAGGAGT TCAAGACCAG CCTGGCCAAC ATGATGAAAC CCTGCCTCTA CTAAAAATAC 2849
AAAAATTAGC AGGGCATGGT GGTGCATGCC TGCAACCCTA GCTACTCAGG AGGCTGAGAC 2909
AGGAGAATCT CTTGAACCCT CGAGGCGGAG GTTGTGGTGA GCTGAGATCC CTCTACTGCA 2969
CTCCAGCCTG GGTGACAGAG ATGAGACTCC GTCCCTGCCG CCGCCCCCGC CTTCCCCCCC 3029
AAAAAGATTC TTCTTCATGC AGAACATACG GCAGTCAACA AAGGGAGACC TGGGTCCAGG 3089
TGTCCAAGTC ACTTATTTCG AGTAAATTAG CAATGAAAGA ATGCCATGGA ATCCCTGCCC 3149
AAATACCTCT GCTTATGATA TTGTAGAATT TGATATAGAG TTGTATCCCA TTTAAGGAGT 3209
AGGATGTAGT AGGAAAGTAC TAAAAACAAA CACACAAACA GAAAACCCTC TTTGCTTTGT 3269
AAGGTGGTTC CTAAGATAAT GTCAGTGCAA TGCTGGAAAT AATATTTAAT ATGTGAAGGT 3329
TTTAGGCTGT GTTTTCCCCT CCTGTTCTTT TTTTCTGCCA GCCCTTTGTC ATTTTTGCAG 3389
GTCAATGAAT CATGTAGAAA GAGACAGGAG ATGAAACTAG AACCAGTCCA TTTTGCCCCT 3449
TTTTTTATTT TCTGGTTTTG GTAAAAGATA CAATGAGGTA GGAGGTTGAG ATTTATAAAT 3509
GAAGTTTAAT AAGTTTCTGT AGCTTTGATT TTTCTCTTTC ATATTTGTTA TCTTGCATAA 3569
GCCAGAATTG GCCTGTAAAA TCTACATATG GATATTGAAG TCTAAATCTG TTCAACTAGC 3629
TTACACTAGA TGGAGATATT TTCATATTCA GATACACTGG AATGTATGAT CTAGCCATGC 3689
GTAATATAGT CAAGTGTTTG AAGGTATTTA TTTTTAATAG CGTCTTTAGT TGTGGACTGG 3749
TTCAAGTTTT TCTGCCAATG ATTTCTTCAA ATTTATCAAA TATTTTTCCA TCATGAAGTA 3809
AAATGCCCTT GCAGTCACCC TTCCTGAAGT TTGAACGACT CTGCTGTTTT AAACAGTTTA 3869
AGCAAATGGT ATATCATCTT CCGTTTACTA TGTAGCTTAA CTGCAGGCTT ACGCTTTTGA 3929
GTCAGCGGCC AACTTTATTG CCACCTTCAA AAGTTTATTA TAATGTTGTA AATTTTTACT 3989
TCTCAAGGTT AGCATACTTA GGAGTTGCTT CACAATTAGG ATTCAGGAAA GAAAGAACTT 4049
CAGTAGGAAC TGATTGGAAT TTAATGATGC AGCATTCAAT GGGTACTAAT TTCAAAGAAT 4109
GATATTACAG CAGACACACA GCAGTTATCT TGATTTTCTA GGAATAATTG TATGAAGAAT 4169
ATGGCTGACA ACACGGCCTT ACTGCCACTC AGCGGAGGCT GGACTAATGA ACACCCTACC 4229
CTTCTTTCCT TTCCTCTCAC ATTTCATGAG CGTTTTGTAG GTAACGAGAA AATTGACTTG 4289
CATTTGCATT ACAAGGAGGA GAAACTGGCA AAGGGGATGA TGGTGGAAGT TTTGTTCTGT 4349
CTAATGAAGT GAAAAATGAA AATGCTAGAG TTTTGTGCAA CATAATAGTA GCAGTAAAAA 4409
CCAAGTGAAA AGTCTTTCCA AAACTGTGTT AAGAGGGCAT CTGCTGGGAA ACGATTTGAG 4469
GAGAAGGTAC TAAATTGCTT GGTATTTTCC GTAG GA ACC CCA GAG CGA AAT ACA GA 4523
                                    Gly Thr Pro Glu Arg Asn Thr
                                                               115

GTT TGC AAA AGA TGT CCA GAT GGG TTC TTC TCA AAT GAG ACG TCA TCT GA 4571
Val Cys Lys Arg Cys Pro Asp Gly Phe Phe Ser Asn Glu Thr Ser Ser
120           125               130               135

AAA GCA CCC TGT AGA AAA CAC ACA AAT TGC AGT GTC TTT GGT CTC CTG GA 4619
Lys Ala Pro Cys Arg Lys His Thr Asn Cys Ser Val Phe Gly Leu Leu
          140               145               150

CTA ACT CAG AAA GGA AAT GCA ACA CAC GAC AAC ATA TGT TCC GGA AAC GA 4667
Leu Thr Gln Lys Gly Asn Ala Thr His Asp Asn Ile Cys Ser Gly Asn
          155               160               165

AGT GAA TCA ACT CAA AAA TGT GGA ATA G GTAATTACAT TCCAAAATAC GA 4715
Ser Glu Ser Thr Gln Lys Cys Gly Ile
          170               175

GTCTTTGTAC GATTTTGTAG TATCATCTCT CTCTCTGAGT TGAACACAAG GCCTCCAGCC 4775
ACATTCTTGG TCAAACTTAC ATTTTCCCTT TCTTGAATCT TAACCAGCTA AGGCTACTCT 4835
```

/

```
CGATGCATTA CTGCTAAAGC TACCACTCAG AATCTCTCAA AAACTCATCT TCTCACAGAT 4895
AACACCTCAA AGCTTGATTT TCTCTCCTTT CACACTGAAA TCAAATCTTG CCCATAGGCA 4955
AAGGGCAGTG TCAAGTTTGC CACTGAGATG AAATTAGGAG AGTCCAAACT GTAGAATTCA 5015
CGTTGTGTGT TATTACTTTC ACGAATGTCT GTATTATTAA CTAAAGTATA TATTGGCAAC 5075
TAAGAAGCAA AGTGATATAA ACATGATGAC AAATTAGGCC AGGCATGGTG GCTTACTCCT 5135
ATAATCCCAA CATTTTGGGG GGCCAAGGTA GGCAGATCAC TTGAGGTCAG GATTTCAAGA 5195
CCAGCCTGAC CAACATGGTG AAACCTTGTC TCTACTAAAA ATACAAAAAT TAGCTGGGCA 5255
TGGTAGCAGG CACTTCTAGT ACCAGCTACT CAGGGCTGAG GCAGGAGAAT CGCTTGAACC 5315
CAGGAGATGG AGGTTGCAGT GAGCTGAGAT TGTACCACTG CACTCCAGTC TGGGCAACAG 5375
AGCAAGATTT CATCACACAC ACACACACAC ACACACACAC ACACATTAGA AATGTGTACT 5435
TGGCTTTGTT ACCTATGGTA TTAGTGCATC TATTGCATGG AACTTCCAAG CTACTCTGGT 5495
TGTGTTAAGC TCTTCATTGG GTACAGGTCA CTAGTATTAA GTTCAGGTTA TTCGGATGCA 5555
TTCCACGGTA GTGATGACAA TTCATCAGGC TAGTGTGTGT GTTCACCTTG TCACTCCCAC 5615
CACTAGACTA ATCTCAGACC TTCACTCAAA GACACATTAC ACTAAAGATG ATTTGCTTTT 5675
TTGTGTTTAA TCAAGCAATG GTATAAACCA GCTTGACTCT CCCCAAACAG TTTTTCGTAC 5735
TACAAAGAAG TTTATGAAGC AGAGAAATGT GAATTGATAT ATATATGAGA TTCTAACCCA 5795
GTTCCAGCAT TGTTTCATTG TGTAATTGAA ATCATAGACA AGCCATTTTA GCCTTTGCTT 5855
TCTTATCTAA AAAAAAAAAA AAAAAAATGA AGGAAGGGGT ATTAAAAGGA GTGATCAAAT 5915
TTTAACATTC TCTTTAATTA ATTCATTTTT AATTTTACTT TTTTTCATTT ATTGTGCACT 5975
TACTATGTGG TACTGTGCTA TAGAGGCTTT AACATTATA AAAACACTGT GAAAGTTGCT 6035
TCAGATGAAT ATAGGTAGTA GAACGGCAGA ACTAGTATTC AAAGCCAGGT CTGATGAATC 6095
CAAAAACAAA CACCCATTAC TCCCATTTTC TGGGACATAC TTACTCTACC CAGATGCTCT 6155
GGGCTTTGTA ATGCCTATGT AAATAACATA GTTTTATGTT TGGTTATTTT CCTATGTAAT 6215
GTCTACTTAT ATATCTGTAT CTATCTCTTG CTTTGTTTCC AAAGGTAAAC TATGTGTCTA 6275
AATGTGGGCA AAAAATAACA CACTATTCCA AATTACTGTT CAAATTCCTT TAAGTCAGTG 6335
ATAATTATTT GTTTTGACAT TAATCATGAA GTTCCCTGTG GGTACTAGGT AAACCTTTAA 6395
TAGAATGTTA ATGTTTGTAT TCATTATAAG AATTTTTGGC TGTTACTTAT TTACAACAAT 6455
ATTTCACTCT AATTAGACAT TTACTAAACT TTCTCTTGAA AACAATGCCC AAAAAAGAAC 6515
ATTAGAAGAC ACGTAAGCTC AGTTGGTCTC TGCCACTAAG ACCAGCCAAC AGAAGCTTGA 6575
TTTTATTCAA ACTTTGCATT TTAGCATATT TTATCTTGGA AAATTCAATT GTGTTGGTTT 6635
TTTGTTTTTG TTTGTATTGA ATAGACTCTC AGAAATCCAA TTGTTGAGTA AATCTTCTGG 6695
GTTTTCTAAC CTTTCTTTAG AT GTT ACC CTG TGT GAG GAG GCA TTC TTC AG   6747
                          Asp Val Thr Leu Cys Glu Glu Ala Phe Phe Arg
                          180                              185
```

```
TTT GCT GTT CCT ACA AAG TTT ACG CCT AAC TGG CTT AGT GTC TTG GTA   6795
Phe Ala Val Pro Thr Lys Phe Thr Pro Asn Trp Leu Ser Val Leu Val
          190              195              200
```

```
GAC AAT TTG CCT GGC ACC AAA GTA AAC GCA GAG AGT GTA GAG AGG ATA   6843
Asp Asn Leu Pro Gly Thr Lys Val Asn Ala Glu Ser Val Glu Arg Ile
      205              210              215
```

```
AAA CGG CAA CAC AGC TCA CAA GAA CAG ACT TTC CAG CTG CTG AAG TTA   6891
Lys Arg Gln His Ser Ser Gln Glu Gln Thr Phe Gln Leu Leu Lys Leu
220              225              230              235
```

```
TGG AAA CAT CAA AAC AAA GAC CAA GAT ATA GTC AAG AAG ATC ATC CAA G 6940
Trp Lys His Gln Asn Lys Asp Gln Asp Ile Val Lys Lys Ile Ile Gln
        240              245              250
```

```
GTATGATAAT CTAAAATAAA AAGATCAATC AGAAATCAAA GACACCTATT TATCATAAAC 7000
CAGGAACAAG ACTGCATGTA TGTTTAGTTG TGTGGATCTT GTTTCCCTGT TGGAATCATT 7060
GTTGGACTGA AAAAGTTTCC ACCTGATAAT GTAGATGTGA TTCCACAAAC AGTTATACAA 7120
GGTTTTGTTC TCACCCCTGC TCCCCAGTTT CCTTGTAAAG TATGTTGAAC ACTCTAAGAG 7180
AAGAGAAATG CATTTGAAGG CAGGGCTGTA TCTCAGGGAG TCGCTTCCAG ATCCCTTAAC 7240
GCTTCTGTAA GCAGCCCCTC TAGACCACCA AGGAGAAGCT CTATAACCAC TTTGTATCTT 7300
```

```
ACATTGCACC TCTACCAAGA AGCTCTGTTG TATTTACTTG GTAATTCTCT CCAGGTAGGC 7360
TTTTCGTAGC TTACAAATAT GTTCTTATTA ATCCTCATGA TATGGCCTGC ATTAAAATTA 7420
TTTTAATGGC ATATGTTATG AGAATTAATG AGATAAAATC TGAAAAGTGT TTGAGCCTCT 7480
TGTAGGAAAA AGCTAGTTAC AGCAAAATGT TCTCACATCT TATAAGTTTA TATAAAGATT 7540
CTCCTTTAGA AATGGTGTGA GAGAGAAACA GAGAGAGATA GGGAGAGAAG TGTGAAAGAA 7600
TCTGAAGAAA AGGAGTTTCA TCCAGTGTGG ACTGTAAGCT TTACGACACA TGATGGAAAG 7660
AGTTCTGACT TCAGTAAGCA TTGGGAGGAC ATGCTAGAAG AAAAAGGAAG AAGAGTTTCC 7720
ATAATGCAGA CAGGGTCAGT GAGAAATTCA TTCAGGTCCT CACCAGTAGT TAAATGACTG 7780
TATAGTCTTG CACTACCCTA AAAAACTTCA AGTATCTGAA ACCGGGGCAA CAGATTTTAG 7840
GAGACCAACG TCTTTGAGAG CTGATTGCTT TTGCTTATGC AAAGAGTAAA CTTTTATGTT 7900
TTGAGCAAAC CAAAAGTATT CTTTGAACGT ATAATTAGCC CTGAAGCCGA AAGAAAAGAG 7960
AAAATCAGAG ACCGTTAGAA TTGGAAGCAA CCAAATTCCC TATTTTATAA ATGAGGACAT 8020
TTTAACCCAG AAAGATGAAC CGATTTGGCT TAGGGCTCAC AGATACTAAG TGACTCATGT 8080
CATTAATAGA AATGTTAGTT CCTCCCTCTT AGGTTTGTAC CCTAGCTTAT TACTGAAATA 8140
TTCTCTAGGC TGTGTGTCTC CTTTAGTTCC TCGACCTCAT GTCTTTGAGT TTTCAGATAT 8200
CCTCCTCATG GAGGTAGTCC TCTGGTGCTA TGTGTATTCT TTAAAGGCTA GTTACGGCAA 8260
TTAACTTATC AACTAGCGCC TACTAATGAA ACTTTGTATT ACAAAGTAGC TAACTTGAAT 8320
ACTTTCCTTT TTTTCTGAAA TGTTATGGTG GTAATTTCTC AAACTTTTTC TTAGAAAACT 8380
GAGAGTGATG TGTCTTATTT TCTACTGTTA ATTTTCAAAA TTAGGAGCTT CTTCCAAAGT 8440
TTTGTTGGAT GCCAAAAATA TATAGCATAT TATCTTATTA TAACAAAAAA TATTTATCTC 8500
AGTTCTTAGA AATAAATGGT GTCACTTAAC TCCCTCTCAA AGAAAAGGT TATCATTGAA 8560
ATATAATTAT GAAATTCTGC AAGAACCTTT TGCCTCACGC TTGTTTTATG ATGGCATTGG 8620
ATGAATATAA ATGATGTGAA CACTTATCTG GGCTTTTGCT TTATGCAG AT ATT GAC ⬚⬚8676
                                                        Asp Ile Asp
```

```
CTC TGT GAA AAC AGC GTG CAG CGG CAC ATT GGA CAT GCT AAC CTC ACC ⬚⬚8724
Leu Cys Glu Asn Ser Val Gln Arg His Ile Gly His Ala Asn Leu Thr
255⬚⬚⬚⬚⬚⬚⬚⬚⬚260⬚⬚⬚⬚⬚⬚⬚⬚⬚265⬚⬚⬚⬚⬚⬚⬚⬚⬚270
```

```
TTC GAG CAG CTT CGT AGC TTG ATG GAA AGC TTA CCG GGA AAG AAA GTG ⬚⬚8772
Phe Glu Gln Leu Arg Ser Leu Met Glu Ser Leu Pro Gly Lys Lys Val
⬚⬚⬚⬚⬚⬚⬚⬚⬚275⬚⬚⬚⬚⬚⬚⬚⬚⬚280⬚⬚⬚⬚⬚⬚⬚⬚⬚285
```

```
GGA GCA GAA GAC ATT GAA AAA ACA ATA AAG GCA TGC AAA CCC AGT GAC ⬚⬚8820
Gly Ala Glu Asp Ile Glu Lys Thr Ile Lys Ala Cys Lys Pro Ser Asp
⬚⬚⬚⬚⬚⬚290⬚⬚⬚⬚⬚⬚⬚⬚⬚295⬚⬚⬚⬚⬚⬚⬚⬚⬚300
```

```
CAG ATC CTG AAG CTG CTC AGT TTG TGG CGA ATA AAA AAT GGC GAC CAA ⬚⬚8868
Gln Ile Leu Lys Leu Leu Ser Leu Trp Arg Ile Lys Asn Gly Asp Gln
⬚⬚⬚⬚⬚305⬚⬚⬚⬚⬚⬚⬚⬚⬚310⬚⬚⬚⬚⬚⬚⬚⬚⬚315
```

```
GAC ACC TTG AAG GGC CTA ATG CAC GCA CTA AAG CAC TCA AAG ACG TAC ⬚⬚8916
Asp Thr Leu Lys Gly Leu Met His Ala Leu Lys His Ser Lys Thr Tyr
⬚⬚320⬚⬚⬚⬚⬚⬚⬚⬚⬚325⬚⬚⬚⬚⬚⬚⬚⬚⬚330
```

```
CAC TTT CCC AAA ACT GTC ACT CAG AGT CTA AAG AAG ACC ATC AGG TTC ⬚⬚8964
His Phe Pro Lys Thr Val Thr Gln Ser Leu Lys Lys Thr Ile Arg Phe
335⬚⬚⬚⬚⬚⬚⬚⬚⬚340⬚⬚⬚⬚⬚⬚⬚⬚⬚345⬚⬚⬚⬚⬚⬚⬚⬚⬚350
```

```
CTT CAC AGC TTC ACA ATG TAC AAA TTG TAT CAG AAG TTA TTT TTA GAA ⬚⬚9012
Leu His Ser Phe Thr Met Tyr Lys Leu Tyr Gln Lys Leu Phe Leu Glu
⬚⬚⬚⬚⬚⬚⬚⬚⬚355⬚⬚⬚⬚⬚⬚⬚⬚⬚360⬚⬚⬚⬚⬚⬚⬚⬚⬚365
```

```
ATG ATA GGT AAC CAG GTC CAA TCA GTA AAA ATA AGC TGC TTA ⬚⬚⬚⬚⬚⬚9054
Met Ile Gly Asn Gln Val Gln Ser Val Lys Ile Ser Cys Leu
⬚⬚⬚⬚⬚⬚⬚⬚⬚370⬚⬚⬚⬚⬚⬚⬚⬚⬚375⬚⬚⬚⬚⬚⬚⬚⬚⬚380
```

```
TAACTGGAAA TGGCCATTGA GCTGTTTCCT CACAATTGGC GAGATCCCAT GGATGAGTAA 9114
ACTGTTTCTC AGGCACTTGA GGCTTTCAGT GATATCTTTC TCATTACCAG TGACTAATTT 9174
TGCCACAGGG TACTAAAAGA AACTATGATG TGGAGAAAGG ACTAACATCT CCTCCAATAA 9234
ACCCCAAATG GTTAATCCAA CTGTCAGATC TGGATCGTTA TCTACTGACT ATATTTTCCC 9294
TTATTACTGC TTGCAGTAAT TCAACTGGAA ATTAAAAAAA AAAAACTAGA CTCCACTGGG 9354
CCTTACTAAA TATGGGAATG TCTAACTTAA ATAGCTTTGG GATTCCAGCT ATGCTAGAGG 9414
CTTTTATTAG AAAGCCATAT TTTTTTCTGT AAAAGTTACT AATATATCTG TAACACTATT 9474
ACAGTATTGC TATTTATATT CATTCAGATA TAAGATTTGG ACATATTATC ATCCTATAAA 9534
GAAACGGTAT GACTTAATTT TAGAAAGAAA ATTATATTCT GTTTATTATG ACAAATGAAA 9594
GAGAAAATAT ATATTTTTAA TGGAAAGTTT GTAGCATTTT TCTAATAGGT ACTGCCATAT 9654
TTTTCTGTGT GGAGTATTTT TATAATTTTA TCTGTATAAG CTGTAATATC ATTTTATAGA 9714
AAATGCATTA TTTAGTCAAT TGTTTAATGT TGGAAAACAT ATGAAATATA AATTATCTGA 9774
ATATTAGATG CTCTGAGAAA TTGAATGTAC CTTATTTAAA AGATTTTATG GTTTTATAAC 9834
TATATAAATG ACATTATTAA AGTTTTCAAA TTATTTTTTA TTGCTTTCTC TGTTGCTTTT 9894
ATT 9898
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) Length of sequence: 401

(B) Sequence type: amino acid

(C) Strandedness: single stranded

(D) Topology: linear

(ii) Molecular type: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Asn Asn Leu Leu Cys Cys Ala Leu Val Phe Leu Asp Ile Ser
    -20              -15            -10
Ile Lys Trp Thr Thr Gln Glu Thr Phe Pro Pro Lys Tyr Leu His
   -5               1              5
Tyr Asp Glu Glu Thr Ser His Gln Leu Leu Cys Asp Lys Cys Pro
 10               15             20
Pro Gly Thr Tyr Leu Lys Gln His Cys Thr Ala Lys Trp Lys Thr
 25               30             35
Val Cys Ala Pro Cys Pro Asp His Tyr Tyr Thr Asp Ser Trp His
 40               45             50
Thr Ser Asp Glu Cys Leu Tyr Cys Ser Pro Val Cys Lys Glu Leu
 55               60             65
Gln Tyr Val Lys Gln Glu Cys Asn Arg Thr His Asn Arg Val Cys
 70               75             80
Glu Cys Lys Glu Gly Arg Tyr Leu Glu Ile Glu Phe Cys Leu Lys
 85               90             95
His Arg Ser Cys Pro Pro Gly Phe Gly Val Val Gln Ala Gly Thr
 100              105            110
Pro Glu Arg Asn Thr Val Cys Lys Arg Cys Pro Asp Gly Phe Phe
 115              120            125
Ser Asn Glu Thr Ser Ser Lys Ala Pro Cys Arg Lys His Thr Asn
 130              135            140
Cys Ser Val Phe Gly Leu Leu Leu Thr Gln Lys Gly Asn Ala Thr
 145              150            155
His Asp Asn Ile Cys Ser Gly Asn Ser Glu Ser Thr Gln Lys Cys
 160              165            170
Gly Ile Asp Val Thr Leu Cys Glu Glu Ala Phe Phe Arg Phe Ala
 175              180            185
Val Pro Thr Lys Phe Thr Pro Asn Trp Leu Ser Val Leu Val Asp
 190              195            200
Asn Leu Pro Gly Thr Lys Val Asn Ala Glu Ser Val Glu Arg Ile
 205              210            215
Lys Arg Gln His Ser Ser Gln Glu Gln Thr Phe Gln Leu Leu Lys
```

(handwritten in right margin) please ensure all amino acids are from those shown in 1.822(b)(2) of Sequence Rules

07/(0)1,01

```
@220                    225                    230
@Leu Trp Lys His Gln Asn Lys Asp Gln Asp Ile Val Lys Lys Ile
@235                    240                    245
@Ile Gln Asp Ile Asp Leu Cys Glu Asn Ser Val Gln Arg His Ile
@250                    255                    260
@Gly His Ala Asn Leu Thr Phe Glu Gln Leu Arg Ser Leu Met Glu
@265                    270                    275
@Ser Leu Pro Gly Lys Lys Val Gly Ala Glu Asp Ile Glu Lys Thr
@280                    285                    290
@Ile Lys Ala Cys Lys Pro Ser Asp Gln Ile Leu Lys Leu Leu Ser
@295                    300                    305
@Leu Trp Arg Ile Lys Asn Gly Asp Gln Asp Thr Leu Lys Gly Leu
@310                    315                    320
@Met His Ala Leu Lys His Ser Lys Thr Tyr His Phe Pro Lys Thr
@325                    330                    335
@Val Thr Gln Ser Leu Lys Lys Thr Ile Arg Phe Leu His Ser Phe
@340                    345                    350
@Thr Met Tyr Lys Leu Tyr Gln Lys Leu Phe Leu Glu Met Ile Gly
@355                    360                    365
@Asn Gln Val Gln Ser Val Lys Ile Ser Cys Leu
@370                    375                    380
```

Sequence number: 4 (2) INFORMATION FOR SEQ ID NO:4:
(A) Length of sequence: 1206 7(i) SEQUENCE CHARACTERISTICS:
(B) Sequence type: nucleic acid
(C) Strandedness: single stranded
(D) Topology: linear
(ii) Molecular type: cDNA
Sequence: (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
ATGAACAACT TGCTGTGCTG CGCGCTCGTG TTTCTGGACA TCTCCATTAA GTGGACCACC   60
CAGGAAACGT TTCCTCCAAA GTACCTTCAT TATGACGAAG AAACCTCTCA TCAGCTGTTG  120
TGTGACAAAT GTCCTCCTGG TACCTACCTA AAACAACACT GTACAGCAAA GTGGAAGACC  180
GTGTGCGCCC CTTGCCCTGA CCACTACTAC ACAGACAGCT GGCACACCAG TGACGAGTGT  240
CTATACTGCA GCCCCGTGTG CAAGGAGCTG CAGTACGTCA AGCAGGAGTG CAATCGCACC  300
CACAACCGCG TGTGCGAATG CAAGGAAGGG CGCTACCTTG AGATAGAGTT CTGCTTGAAA  360
CATAGGAGCT GCCCTCCTGG ATTTGGAGTG GTGCAAGCTG GAACCCCAGA GCGAAATACA  420
GTTTGCAAAA GATGTCCAGA TGGGTTCTTC TCAAATGAGA CGTCATCTAA AGCACCCTGT  480
AGAAACACA CAAATTGCAG TGTCTTTGGT CTCCTGCTAA CTCAGAAAGG AAATGCAACA  540
CACGACAACA TATGTTCCGG AAACAGTGAA TCAACTCAAA AATGTGGAAT AGATGTTACC  600
CTGTGTGAGG AGGCATTCTT CAGGTTTGCT GTTCCTACAA AGTTTACGCC TAACTGGCTT  660
AGTGTCTTGG TAGACAATTT GCCTGGCACC AAAGTAAACG CAGAGAGTGT AGAGAGGATA  720
AAACGGCAAC ACAGCTCACA AGAACAGACT TTCCAGCTGC TGAAGTTATG GAAACATCAA  780
AACAAAGACC AAGATATAGT CAAGAAGATC ATCCAAGATA TTGACCTCTG TGAAAACAGC  840
GTGCAGCGGC ACATTGGACA TGCTAACCTC ACCTTCGAGC AGCTTCGTAG CTTGATGGAA  900
AGCTTACCGG GAAAGAAAGT GGGAGCAGAA GACATTGAAA AAACAATAAA GGCATGCAAA  960
CCCAGTGACC AGATCCTGAA GCTGCTCAGT TTGTGGCGAA TAAAAAATGG CGACCAAGAC 1020
ACCTTGAAGG GCCTAATGCA CGCACTAAAG CACTCAAAGA CGTACCACTT TCCCAAAACT 1080
GTCACTCAGA GTCTAAAGAA GACCATCAGG TTCCTTCACA GCTTCACAAT GTACAAATTG 1140
TATCAGAAGT TATTTTTAGA AATGATAGGT AACCAGGTCC AATCAGTAAA AATAAGCTGC 1200
TTATAA                                                           1206
```

delete

Please consult sample Sequence Listing (attached) for valid format. Suggestions: consult Sequence Rules

(3) Computer: Apple Macintosh;
(i) Operating System: Macintosh;
(ii) Macintosh File Type: text with line termination
(iii) Line Terminator: Pre-defined by text type file;
(iv) Pagination: Pre-defined by text type file;
(v) End-of-file: Pre-defined by text type file;
(vi) Media: (A) Diskett—3.50 inch, 400 Kb storage;
(B) Diskette—3.50 inch, 800 Kb storage;
(C) Diskette—3.50 inch, 1.4 Mb st rage;
(vii) Print Command: Use PRINT command from any Macintosh Application that processes text files, such as MacWrite or Teach Text;
(4) Magnetic tape: 0.5 inch, up to 2400 feet;
(i) Density: 1600 or 6250 bits per inch, 9 track;
(ii) Format: raw, unblocked;
(iii) Line Terminator: ASCII Carriage Return plus optional ASCII Line Feed;
(iv) Pagination: ASCII Form Feed or Series of Line Terminators;
(v) Print Command (Unix shell version given here as sample response—mt/dev/rmt0; lpr/dev/rmt0):
(g) Computer readable forms that are submitted to the Office will not be returned to the applicant.
(h) All computer readable forms shall have a label permanently affixed thereto on which has been hand printed or typed, a description of the format of the computer readable form as well as the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form and the name and type of computer and operating system which generated the files on the computer readable form. If all of this information cannot be printed on a label affixed to the computer readable form, by reason of size or otherwise, the label shall include the name of the applicant and the title of the invention and a reference number, and the additional information may be provided on a container for the computer readable form with the name f the applicant, the title of the invention, the reference number and the additional information affixed to the container. If the computer readable form is submitted after the date f filing

under 35 U.S.C. 111, after the date of entry in the national stage und r 35 U.S.C. 371 or after the time of filing, in the United States Receiving Office, an international application under the PCT, the labels mentioned herein must also include the date of the application and the application number, including series code and serial number.

§ 1.825 Amendments to or replacement of sequence listing and computer readable copy thereof.

(a) Any amendment to the paper copy of the "Sequence Listing" (§ 1.821(c)) must be made by the submission of substitute sheets. Amendments must be accompanied by a statement that indicates support for the amendment in the application, as filed, and a statement that the substitute sheets include no new matter. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(b) Any amendment to the paper copy of the "Sequence Listing," in accordance with paragraph (a) of this section, must be accompanied by a substitute copy of the computer readable form (§ 1.821(e)) including all previously submitted data with the amendment incorporated therein, accompanied by a statement that the copy in computer readable form is the same as the substitute copy of the "Sequence Listing." Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(c) Any appropriate amendments to the "Sequence Listing" in a patent, e.g., by reason of reissue or certificate of correction, must comply with the requirements of paragraphs (a) and (b) of this section.

(d) If, upon receipt, the computer readable form is found to be damaged or unreadable, applicant must provide, within such time as set by the Commissioner, a substitute copy of the data in computer readable form accompanied by a statement that the substitute data is identical to that originally filed. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

Appendix A—Sample Sequence Listing
(1) GENERAL INFORMATION:

(I) APPLICANT: Doe, Joan X, Doe, John Q
(II) TITLE OF INVENTION: Isolation and Characterization of a Gene Encoding a Protease from *Paramecium* sp.
(III) NUMBER OF SEQUENCES: 2
(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Smith and Jones
(B) STREET: 123 Main Street
(C) CITY: Smalltown
(D) STATE: Anystate
(E) COUNTRY: USA ·
(F) ZIP: 12345
(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb storage
(B) COMPUTER: Apple Macintosh
(C) OPERATING SYSTEM: Mcintosh 5.0
(D) SOFTWARE: MacWrite
(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 09/999,999
(B) FILING DATE: 28-FEB-1989
(C) CLASSIFICATION: 999/99
(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US88/99999
(B) FILING DATE: 01-MAR-1988
(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Smith, John A.
(B) REGISTRATION NUMBER: 00001
(C) REFERENCE/DOCKET NUMBER: 01-0001
(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (909) 999-0001
(B) TELEFAX: (909) 999-0002
(2) INFORMATION FOR SEQ ID NO: 1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 954 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Paramecium sp
(C) INDIVIDUAL/ISOLATE: XYZ2
(G) CELL TYPE: unicellular organism
(vii) IMMEDIATE SOURCE:
(A) LIBRARY: genomic
(B) CLONE: Para-XYZ2/36
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Doe, Joan X, Doe, John Q
(B) TITLE: Isolation and Characterization of a Gene Encoding a Protease from *Paramecium* sp.
(C) JOURNAL: Fictional Genes
(D) VOLUME: I
(E) ISSUE: 1
(F) PAGES: 1-20
(G) DATE: 02-MAR-1988
(K) RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 954

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATCGGGATAG TACTGGTCAA GACCGGTGGA CACCGGTTAA CCCCGGTTAA GTACCGGTTA   60

TAGGCCATTT CAGGCCAAAT GTGCCCAACT ACGCCAATTG TTTTGCCAAC GGCCAACGTT  120

ACGTTCGTAC GCACGTATGT ACCTAGGTAC TTACGGACGT GACTACGGAC ACTTCCGTAC  180

GTACGTACGT TTACGTACCC ATCCCAACGT AACCACAGTG TGGTCGCAGT GTCCCAGTGT  240

ACACAGACTG CCAGACATTC TTCACAGACA CCCC ATG ACA CCA CCT GAA CGT CTC  295
                                   Met Thr Pro Pro Glu Arg Leu
                                                          -30

TTC CTC CCA AGG GTG TGT GGC ACC ACC CTA CAC CTC CTC CTT CTG GGG   343
Phe Leu Pro Arg Val Cys Gly Thr Thr Leu His Leu Leu Leu Leu Gly
      -25               -20               -15

CTG CTG CTG GTT CTG CTG CCT GGG GCC CAT GTGAGGCAGC AGGAGAATGG   393
Leu Leu Leu Val Leu Leu Pro Gly Ala His
      -10               -5

GGTGGCTCAG CCAAACCTTG AGCCCTAGAG CCCCCCTCAA CTCTGTTCTC CTAG GGG   450
                                                              Gly

CTC ATG CAT CTT GCC CAC AGC AAC CTC AAA CCT GCT GCT CAC CTC ATT   498
Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His Leu Ile
 1               5                  10                  15

GTAAACATCC ACCTGACCTC CCAGACATGT CCCCACCAGC TCTCCTCCTA CCCCTGCCTC  558

AGGAACCCAA GCATCCACCC CTCTCCCCCA ACTTCCCCCA CGCTAAAAAA AACAGAGGGA  618

GCCCACTCCT ATGCCTCCCC CTGCCATCCC CCAGGAACTC AGTTGTTCAG TGCCCACTTC  678

TAC CCC AGC AAG CAG AAC TCA CTG CTC TGG AGA GCA AAC ACG GAC CGT   726
Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr Asp Arg
          20                  25                  30

GCC TTC CTC CAG GAT GGT TTC TCC TTG AGC AAC AAT TCT CTC CTG GTC   774
Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu Val
          35                  40                  45            —

TAGAAAAAAT AATTGATTTC AAGACCTTCT CCCCATTCTG CCTCCATTCT GACCATTTCA  834

GGGGTCGTCA CCACCTCTCC TTTGGCCATT CCAACAGCTC AAGTCTTCCC TGATCAAGTC  894

ACCGGAGCTT TCAAAGAAGG AATTCTAGGC ATCCCAGGGG ACCCACACCT CCCTGAACCA  954
```

(2) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 82 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(ix) FEATURE:
   (A) NAME/KEY: signal sequence
   (B) LOCATION: -34 to -1

   (C) IDENTIFICATION METHOD: similarity
      to other signal sequences, hydrophobic
   (D) OTHER INFORMATION: expresses
      protease
(x) PUBLICATION INFORMATION:
   (A) AUTHORS: Doe, Joan X, Doe, John Q
   (B) TITLE: Isolation and Characterization
      of a Gene Encoding a Protease from
      Paramecium sp.

   (C) JOURNAL: Fictional Genes
   (D) VOLUME: I
   (E) ISSUE: 1
   (F) PAGES: 1–20
   (G) DATE: 02-MAR-1988
   (K) RELEVANT RESIDUES IN SEQ ID NO:
      2: FROM -34 TO 48

BILLING CODE 3510-16-M

Here's where sequence 2 starts (after the sequence data of SEQ ID NO:1:)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
              -30             -25             -20

Leu His Leu Leu Leu Leu Gly Leu Leu Leu Val Leu Leu Pro Gly Ala
          -15             -10             -5

His Gly Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His
          1               5               10

Leu Ile Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr
15              20              25              30

Asp Arg Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu
              35              40              45

Leu Val
```

BILLING CODE 3510-16-C